Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4534–4538, May 1992
Biochemistry

Cloning of the yeast *FAS3* gene and primary structure of yeast acetyl-CoA carboxylase

(acetyl-CoA carboxylase gene/biotin carboxylase/transcarboxylase)

WALID AL-FEEL, SUBRAHMANYAM S. CHIRALA, AND SALIH J. WAKIL*

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Contributed by Salih J. Wakil, December 24, 1991

ABSTRACT **We have isolated and determined the nucleotide sequence of the yeast *FAS3* gene, which encodes acetyl-CoA carboxylase (EC 6.4.1.2). The sequence has an open reading frame of 6711 bases coding for a protein of 2237 amino acids with a calculated molecular weight of 250,593. The presence of the unique biotin-binding site, Met-Lys-Met, and the known CNBr peptide and COOH-terminal sequences confirmed the nucleotide-derived amino acid sequence. The yeast, chicken, and rat carboxylases have an overall sequence identity of 34%, suggesting that the eukaryotic carboxylase evolved from a single ancestral gene. The amino acid sequences of yeast fatty acid synthase subunits are least homologous with the animal synthase sequences, whereas carboxylase sequences are highly conserved. The sequences of the ATP, HCO_3^-, and CoA binding sites of the carboxylases are also well conserved (≈50% identical). The sequences surrounding the biotin binding site are poorly conserved, suggesting that this sequence may not be critical as long as the biotin is available for carboxylase reactions. On the basis of this sequence identity, we have defined the putative biotin carboxylase and transcarboxylase domains.**

Acetyl-CoA carboxylase (ACC; EC 6.4.1.2) catalyzes the committed step in fatty acid biosynthesis, yielding malonyl-CoA, the donor of the two-carbon units for the synthesis of long-chain fatty acids (1). In prokaryotes, the enzyme consists of three readily dissociated proteins, the biotin carboxyl carrier protein (BCCP), the biotin carboxylase, and the transcarboxylase (2), whereas in higher and lower eukaryotic cells, these proteins are part of a single multifunctional polypeptide derived from the expression of a single gene that, presumably, evolved by the fusion of individual genes. Recently, the cDNAs coding for the rat (3) and chicken (4) ACC were cloned and sequenced. The sites for the biotin attachments in both carboxylases, which are conserved as in all other biotin-containing enzymes, were readily identified. Putative domains for the two catalytic functions were assigned (3, 4) and sites for phosphorylation were located (5).

ACC has been isolated from *Saccharomyces cerevisiae* (6) and *Candida lipolytica* (7) and shown to contain single-subunit proteins of molecular weights 189,000 and 230,000, respectively. To understand the relationship between structure and function of the yeast ACC and to utilize the potential of genetic manipulations in yeast, we have undertaken a systematic analysis of the enzyme. In this report, we describe the isolation and nucleotide sequence of the *FAS3* gene,† which encodes yeast ACC, and compare the deduced amino acid sequence with sequences of the rat and chicken enzymes to determine their evolutionary relationship.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Preparation of ACC and Its Antibodies. Yeast ACC from extracts of baker's yeast was purified to a state of homogeneity. Cell-free extracts were prepared, as described previously (8), and the enzyme was isolated by ammonium sulfate fractionation (0–28% saturation), polyethylene glycol 8000 (0–6%) precipitation, and avidin-Sepharose affinity chromatography (9). The carboxylase preparation had specific activity of 2.5 units/mg protein when assayed by [^{14}C]bicarbonate incorporation into malonyl-CoA, as described (9). Antibodies against the purified ACC were raised in rabbits, and affinity-purified anti-ACC antibodies were prepared (10).

Isolation and Sequencing of the *FAS3* Gene. The yeast genomic DNA libraries in λgt11 and EMBL3a vectors were provided by M. Snyder (Yale University) (11). The λgt11 library was screened with anti-ACC antibodies by following standard procedures (11, 12), and the EMBL3a library was screened with radioactive DNA probes as described (13). The yeast strain SEY2102 was grown in appropriate media and total RNA was isolated for Northern analysis (14).

DNA sequencing was performed by using both the dideoxynucleotide termination method (15), as described previously (14), and an automated DNA sequencer (Applied Biosystems model 370A), according to manufacturer's recommendations. All the restriction enzymes and other chemicals were purchased from commercial and standard sources (9, 14).

RESULTS AND DISCUSSION

Isolation and Expression of λgt11acc. The native yeast ACC is a tetramer of identical subunits each having an estimated molecular weight of 250,000 (Fig. 1). Affinity-purified anti-ACC antibodies were used to screen a yeast λgt11 genomic DNA expression library (11, 12). A positive clone, λgt11acc, was isolated and shown to contain a 3.0-kilobase-pair (kbp) fragment of the putative ACC genomic DNA. Initial verification came from DNA sequence analysis of an *Sst* I–*Sst* I fragment from λgt11acc (see Fig. 3). The nucleotide-derived amino acid sequence showed a high degree of sequence identity with animal ACC DNA sequences, and, on the basis of this homology, it was concluded that the λgt11acc clone contained the portion of the *FAS3* gene coding for the COOH terminus of yeast ACC. The identity of this clone was verified further by immunoblotting (16) of protein lysates from λgt11acc/Y1089 lysogens (10). As shown in Fig. 1, a fusion protein (M_r ≈ 180,000) produced by the λgt11acc recombinant phage reacted with the anti-ACC antibodies. On the basis of the sizes of the fusion protein and the β-galactosidase, the carboxylase gene fragment in λgt11acc coded for

Abbreviations: ACC, acetyl-CoA carboxylase; BCCP, biotin carboxyl carrier protein.

*To whom reprints should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M92156).

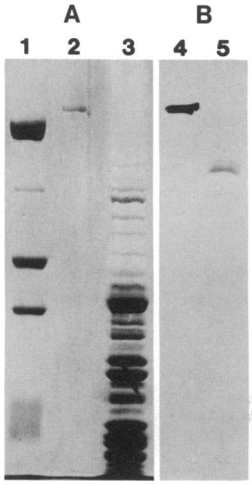

FIG. 1. SDS/PAGE and Western blot analyses of yeast ACC and the β-galactosidase-fused protein produced by λgt11acc. (*A*) Coomassie blue-stained 5% polyacrylamide denaturing gel. Lane 1, protein molecular weight standards (myosin, 200,000; β-galactosidase, 116,000; and phosphorylase b, 97,000); lane 2, yeast ACC (5 μg); and lane 3, cell lysate obtained from λgt11acc lysogen. (*B*) Western blot analysis of a similar gel with anti-yeast ACC antibodies. Lane 4, yeast ACC; and lane 5, λgt11acc cell lysate.

a protein of about M_r 66,000. However, the sizes of the expressed protein and the cloned genomic DNA fragment were smaller than the carboxylase subunit protein of M_r 250,000 and the expected size of about 7.0 kbp of the ACC gene, respectively. To determine if λgt11acc hybridizes to a high molecular weight RNA, Northern analysis (17) was performed using total yeast RNA (13) in conjunction with a ^{32}P-labeled *Eco*RI fragment from λgt11acc. As shown in Fig. 2, the 0.6-kbp yeast DNA fragment isolated from λgt11acc hybridized to an mRNA of 7.5 kilobases (kb), which is larger than 6.6-kb *FAS1* mRNA (14) and is consistent with the sizes of yeast ACC and β subunit of yeast fatty acid synthase.

A 1.9-kbp *Kpn* I restriction fragment from λgt11acc (Fig. 3) was used as a probe for screening an EMBL3a yeast genomic library according to standard procedures (13). A clone, EMBL-ACC, was isolated and shown to contain a yeast DNA insert of about 14 kbp. Southern analysis (data not shown) indicated that clone EMBL-ACC contained the entire 7.3-kbp coding region together with flanking noncoding regions of the yeast ACC gene.

Sequence Analysis of the ACC Gene. The restriction map and sequence strategy used in the structural analyses of the ACC coding region are outlined in Fig. 3. The DNA was sequenced by using standard procedures (18). More than 95% of the sequence was confirmed by sequencing both strands. The remaining sequences were confirmed by sequencing the fragments more than twice in the same direction.

The nucleotide sequence of the DNA encoding the ACC and the derived amino acid sequence are shown in Fig. 4. Starting with the first ATG (Met) codon at nucleotide 1, the nucleotide sequence has an open reading frame of 6711 bases coding for a protein of 2237 amino acids having a molecular weight of 250,593. There are no introns in the entire sequence, since the highly conserved intron-specific sequence TACTAAC is not present. In all three reading frames the nucleotide sequence upstream of the putative initiation codon

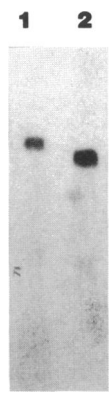

FIG. 2. Northern blot analyses of total yeast RNA (10 μg). Lane 1, yeast RNA hybridized with 0.6-kbp *Eco*RI fragment prepared from λgt11acc; and lane 2, RNA hybridized with a 2.8-kbp Hind-Bam DNA fragment obtained from *FAS1* in YEP33F1 (14).

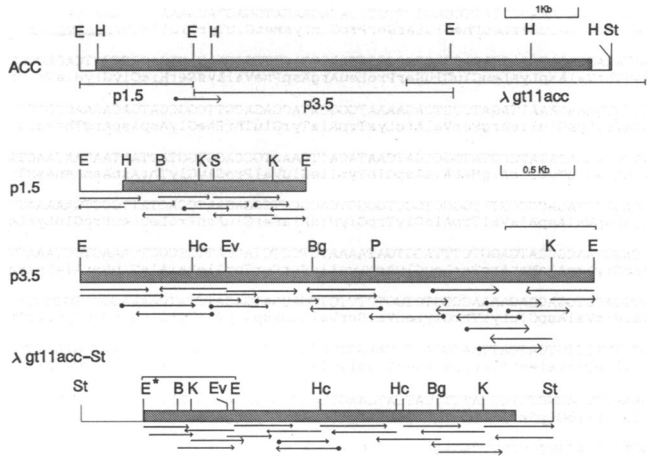

FIG. 3. Restriction map and sequence determination strategy of yeast ACC genomic DNA (ACC) cloned in EMBL3a. *Eco*RI fragments p1.5 and p3.5 were subcloned in the pBluescript vector (Stratagene). The *Sst* I fragment from the yeast DNA in λgt11acc was isolated and subcloned in PUC 19. Only the restriction sites used for sequencing are indicated. Each fragment was sequenced to the extent of each of the arrows in the direction shown. Arrows with nothing or closed circles at the unpointed ends denote sequencing using internal restriction sites and appropriate oligonucleotide primers, respectively. Hatched bars indicate the coding segment. The bracketed areas in p3.5 and λgt11acc-St are the overlapping regions. Restriction sites: B, *Bam*H I; Bg, *Bgl* II; E, *Eco*RI; Ev, *Eco*RV; H, *Hin*dIII; Hc, *Hin*cII; K, *Kpn* I; P, *Pst* I; St, *Sst* I; E*, *Eco*RI linker not in gene.

terminated (data not shown), indicating that the coding sequence can start only with this Met codon or one of the internal (downstream) Met codons. When the rat and yeast amino acid sequences were aligned, we found that the linear homology starts from amino acid residues 126 and 68 in the rat and yeast, respectively, allowing only the Met codons present in the first 68 amino acids as likely candidates for translation initiation sites. Thus, the first Met codon and/or a Met a residue 14, both of which have a purine at the −3 position (19), are probable translation initiation sites, and for now, we are considering that it is the first ATG. The open reading frame ends at residue 2237 with Leu-Lys, consistent with the COOH-terminal sequence of yeast ACC, as reported by Lynen (20). We have sequenced a peptide isolated from yeast ACC after cleavage with CNBr that exactly matches the sequence from amino acid residues 2019 to 2026 (underlined sequence in Fig. 4). The conserved biotin-binding site, Met-Lys-Met, is located between amino acid residues 734 and 736.

Protein Structure and Functional Domains. ACCs are biotin enzymes that are highly conserved, so much so that antibodies prepared against rat, chicken, and yeast enzymes cross-react with each other (unpublished results). The rat (3) and chicken (4) ACCs, which consist of 2345 and 2321 amino acids, respectively, are highly homologous (90% identical), despite the evolutionary diversity, and, hence, little information can be derived by comparing the two sequences. The overall sequence similarity between the rat enzyme and the yeast carboxylase is about 34%, which is significant considering their evolutionary divergence. Further, along these sequences are several stretches of various lengths that are about 80% conserved, with some segments even 100%.

Closer examination of the amino acid sequences of rat and yeast enzymes showed other variable regions as well as regions of high similarity, as illustrated in Fig. 5. For one, the amino acid sequence of the yeast ACC is shorter than that of the rat enzyme by 108 residues. The shortage occurs near the NH_2 terminus, where there are stretches of 50 and 8 amino acids missing, and the COOH terminus, where the enzyme is 20 residues shorter than the rat ACC. There are also notable

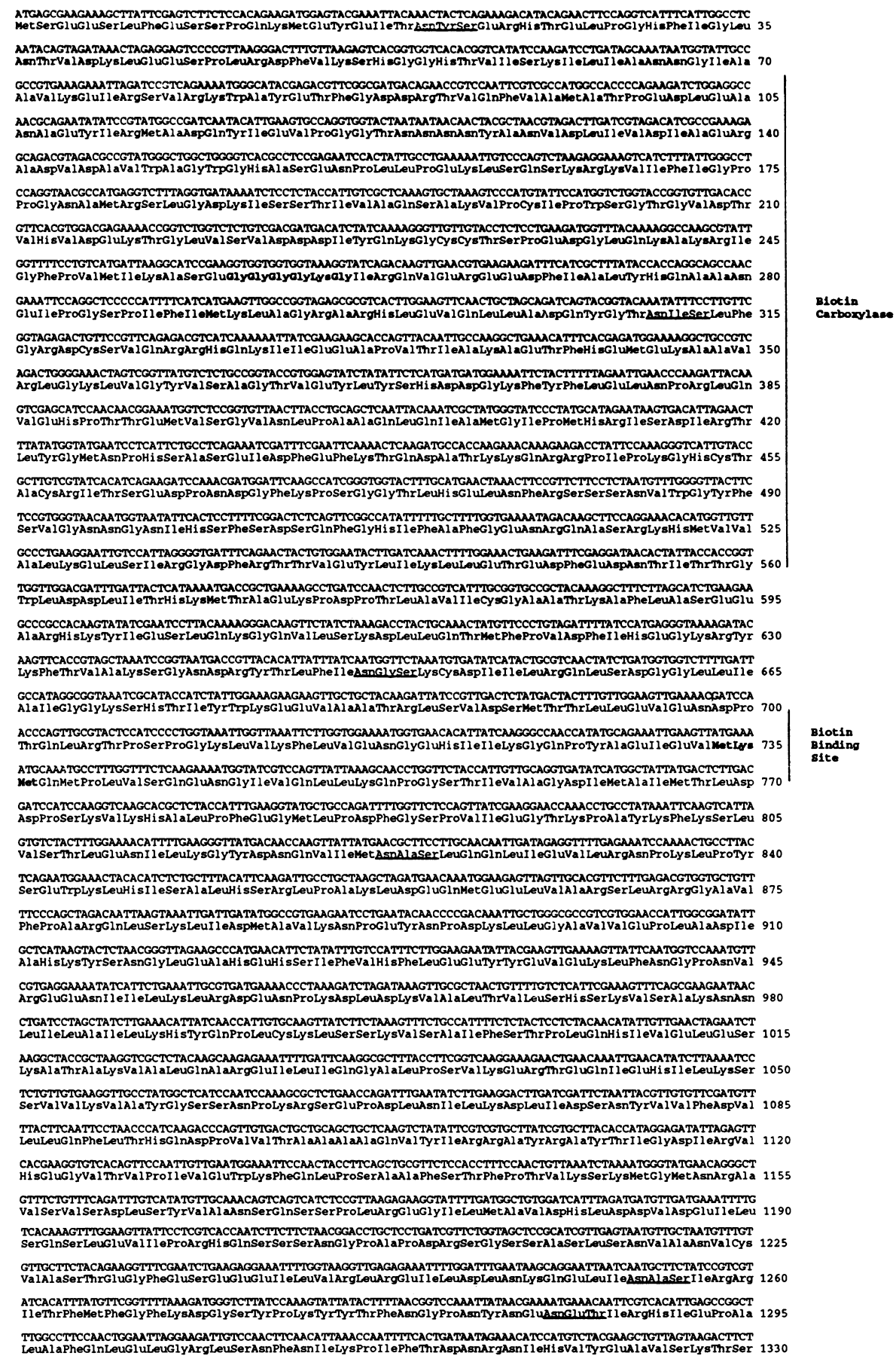

FIG. 4. (*Figure continues on the opposite page.*)

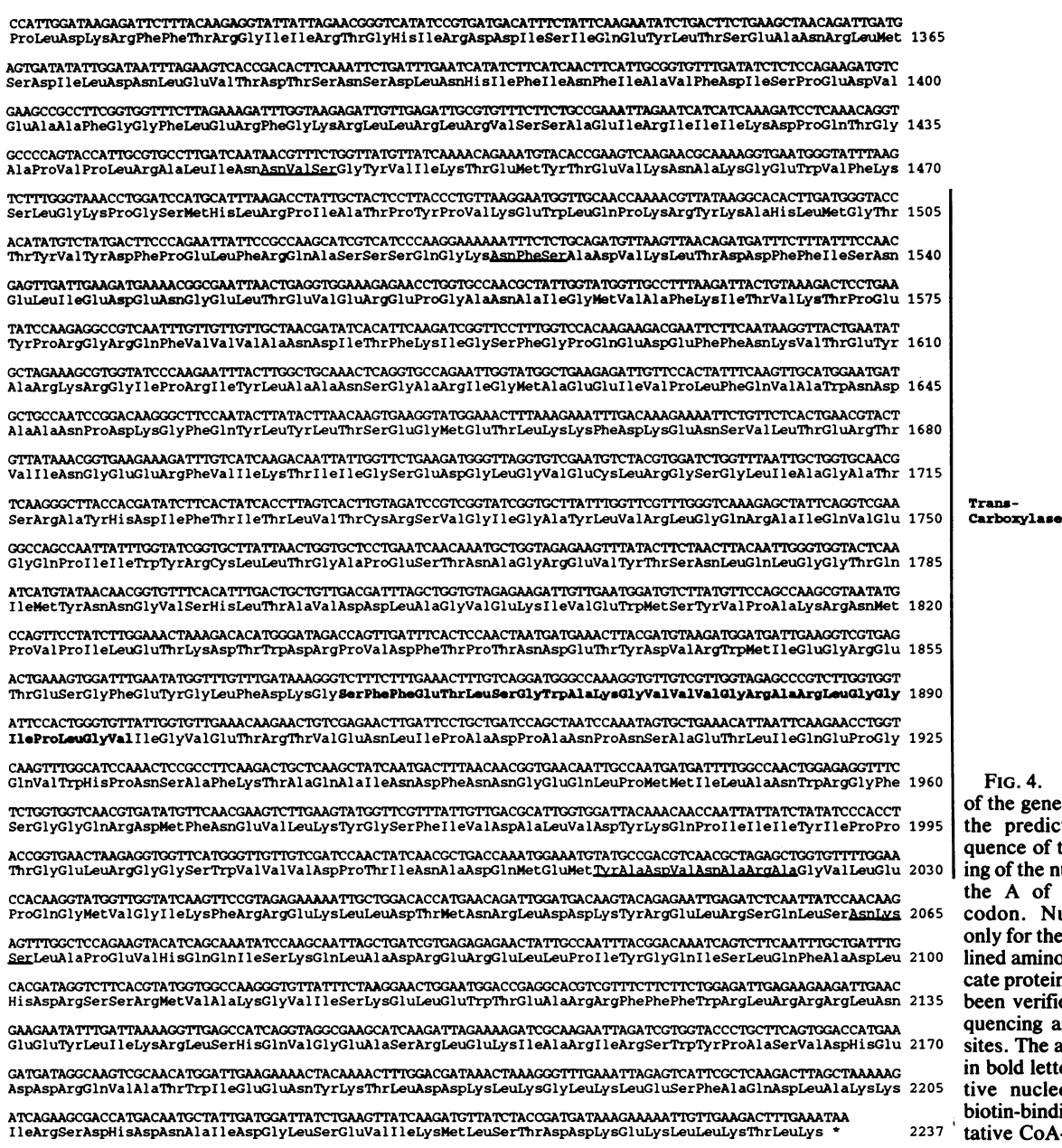

FIG. 4. Nucleotide sequence of the gene coding for ACC and the predicted amino acid sequence of the protein. Numbering of the nucleotides starts with the A of the first ATG Met codon. Numbering is shown only for the amino acids. Underlined amino acid sequences indicate protein sequences that have been verified by amino acid sequencing and the glycosylation sites. The amino acid sequences in bold letters indicate the putative nucleotide-binding motif, biotin-binding site, and the putative CoA-binding site.

internal deletions of 17 and 18 amino acids near residues 1249 and 1328, respectively. Other shorter deletions throughout both yeast and rat ACC make up the overall difference. A surprising lack of homology occurs between the sequence of the first 100 amino acids of rat, the region of the rat enzyme that can be phosphorylated *in vitro* by six protein kinases (5), and that of the yeast enzyme. In rat ACC, phosphorylation of Ser-77 and Ser-79 by cAMP-dependent and AMP-activated protein kinases decreases carboxylase activity (lower V_{max}) and increases citrate concentration required for half-maximal activation (5). Although there are several serine residues near the NH_2-terminal region of yeast ACC, whether or not one or more of these serine residues are phosphorylated remains to be determined. Other proposed phosphorylation sites on the rat ACC are Ser-1200, phosphorylated by both cAMP and AMP-dependent protein kinases, and Ser-1215, phosphorylated by the AMP-activated kinase (5). Phosphorylation at these sites may be important in the regulation of ACC activity (21). Similarly, the yeast ACC has serine residues in the equivalent peptide region (residues 1200–1220), and we suspect that phosphorylation of a serine residue in this region may result in decreased activity (K. V. Venkatachalam, W.-Y. Huang, and S.J.W., unpublished results).

The polypeptide segments of the yeast ACC that exhibit high amino acid sequence similarity to those of rat enzyme vary in length and occur in three subdomains along the protein (Fig. 5). The first subdomain, which spans the region near the NH_2 terminus (residues 100–600) and exhibits about 50% identity to corresponding sequences of rat enzyme, could be the putative biotin carboxylase component of yeast ACC because of its high sequence similarity to corresponding amino acid sequences of known biotin enzymes such as yeast pyruvate carboxylase (residues 162–355), and the α subunit of human propionyl-CoA carboxylase (residues 177–375). Moreover, within the yeast ACC subdomain, the region between amino acid residues 235 and 392 is highly homologous to a corresponding segment of rat ACC and may contain the ATP and HCO_3^- binding sites. This conclusion was based, in part, on the presence of the Gly-rich motif of Gly-Xaa-Gly-Xaa-Xaa-Gly or Gly-Xaa-Xaa-Gly-Xaa-Gly, which have been suggested as the nucleotide-binding motif for yeast and rat carbamoyl-phosphate synthases and rat ACC (3, 22). Further, Hamada *et al.* (23) have proposed that the sequence Tyr-Gly-Tyr-Thr-His-Leu-Ser-Thr-Gly in rabbit muscle myokinase (residues 32–40) is involved in the binding of MgATP. Similar sequences are found within this subdomain

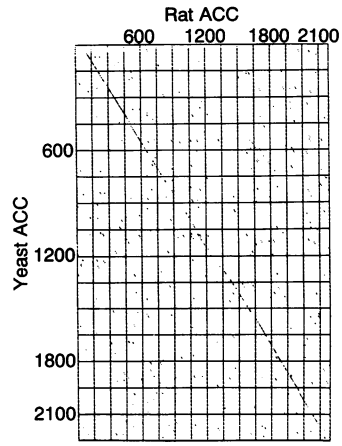

FIG. 5. Dot-matrix plot of rat vs. yeast amino acid sequences of ACC.

of yeast ACC (residues 366–373), Tyr-Leu-Tyr-Ser-His-Asp-Asp-Gly, and in rat ACC (residues 424–432), Tyr-Leu-Tyr-Ser-Gln-Asp-Asp-Ser; however, their significance in nucleotide binding remains to be determined.

The conserved biotin-binding site Met-Lys-Met is located at residues 734–736 in yeast ACC, which represents the region of biotin carboxyl carrier protein, the second subdomain. Comparison of amino acid sequences near the biotin-binding sites with all known biotin-containing enzymes showed that (*i*) the Met-Lys-Met sequences of yeast and animal ACCs are preceded by Val instead of Ala, as in other carboxylases, and (*ii*) these sequences are located closer to the NH_2 termini of the molecules, whereas in other biotin-containing enzymes this sequence occurs near the COOH termini. It has been suggested that this positioning of the biotin-binding site may increase the efficiency of the biotinylation of the proteins (24). In all the biotin-containing enzymes, the biocytin residue is located 25–29 amino acids downstream from a short amino acid sequence flanked by two Pro residues [-Pro-$(Xaa)_n$-Pro-), which might act as a hinge to permit the biotin-containing arm to move between the carboxyl donor and acceptor sites (2). Yeast, rat, and chicken ACCs also contain similar sequences (Fig. 4).

Despite the conservation of biotin-binding sites -Met-Lys-Met- among all biotin-containing proteins, the amino acid sequences surrounding these sites are divergent. Indeed, the sequence of yeast ACC between residues 600 and 1700 is least homologous (26%) with that of the rat ACC (Fig. 5). The lack of conservation of the sequences in these regions of the enzymes may pertain to the assembly of the enzyme subunits into polymer forms that make up the enzymatically active carboxylases. Also, this lack of conservation suggests that the amino acid sequences within BCCP domains of the carboxylases may be involved only in providing a scaffold for the critical regions of the structure to function. In this regard, the biotin in the BCCP domain may be akin to the 4′-phosphopantetheine prosthetic group of the acyl carrier protein domain of the fatty acid synthase, where the quaternary structure of the protein and the need for the pantetheine-bound fatty acyl intermediates to interact with the various catalytic domains of the synthase play a crucial role in the overall activity of the enzymes.

The third subdomain of the yeast ACC, residues 1700–2100, is highly homologous (60%) to the corresponding segment of the rat ACC, residues 1650–2200 (3). Within these subdomains, sequences between amino acid residues 1870 and 1890 are highly similar to the proposed sequence of the "adenine recognition loop" of porcine and yeast citrate synthase (25, 26) and to the β subunit of human propionyl-CoA carboxylase (27). Hence, these peptide segments in ACC are likely to be components of the CoA-binding site.

Recently, Bowers and Allred (28) reported that the rat liver ACC is a glycoprotein. In the yeast carboxylase sequence, there are nine sites that can be N-glycosylated, as indicated in Fig. 4. However, the presence of carbohydrates in the carboxylase and the involvement of any of the putative glycosylation sites remain to be investigated.

We thank C. Green, S. Richards, and K. C. Janssen for their technical help, B. Lee Ligon for editorial assistance, and W.-Y. Huang for providing the amino acid sequence of CNBr peptide. This work was supported in part by Grant DMB 8713051 from the National Science Foundation and a grant from the Clayton Foundation for Research to S.J.W. and Grant DK41872 from the National Institutes of Health to S.S.C.

1. Wakil, S. J., Stoops, J. K. & Joshi, V. C. (1983) *Annu. Rev. Biochem.* **52**, 537–579.
2. Wood, H. G. & Barden, R. E. (1977) *Annu. Review. Biochem.* **46**, 385–413.
3. Lopez-Casillas, F., Bai, D.-H., Luo, X., Kong, I.-S., Hermodson, M. S. & Kim, K.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5784–5788.
4. Takai, T., Yokoyama, C., Wada, K. & Tanabe, T. (1988) *J. Biol. Chem.* **263**, 2651–2657.
5. Davies, S. P., Sim, A. T. R. & Hardie, G. H. (1990) *Eur. J. Biochem.* **198**, 183–190.
6. Sumper, M. & Riepertinger, C. (1972) *Eur. J. Biochem.* **29**, 237–248.
7. Mishina, M., Kamiryo, T., Tanaka, A., Fukui, S. & Numa, S. (1976) *Eur. J. Biochem.* **71**, 295–300.
8. Stoops, J. K., Awad, E. S., Arslanian, M. J., Gunsberg, S., Wakil, S. J. & Oliver, R. M. (1978) *J. Biol. Chem.* **253**, 4464–4475.
9. Thampy, K. G. & Wakil, S. J. (1985) *J. Biol. Chem.* **260**, 6318–6323.
10. Shapiro, D. J., Taylor, J. M., McKnight, G. S., Palacios, R., Gonzales, C., Kiely, M. L. & Schimke, R. T. (1974) *J. Biol. Chem.* **249**, 3665–3671.
11. Snyder, M., Elledge, S., Sweetser, D., Young, R. A. & Davis, R. W. (1987) *Methods Enzymol.* **154**, 107–128.
12. Young, R. A. & Davis, R. W. (1984) *Science* **222**, 778–780.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Chirala, S. S., Kuziora, M. A., Spector, D. M. & Wakil, S. J. (1987) *J. Biol. Chem.* **262**, 4231–4240.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
17. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
18. Mohamed, A. H., Chirala, S. S., Mody, N. H., Huang, W.-Y. & Wakil, S. J. (1988) *J. Biol. Chem.* **263**, 12315–12325.
19. Kodzak, M. (1981) *Nucleic Acids. Res.* **9**, 5233–5252.
20. Lynen, F. (1979) *CRC Crit. Rev. Biochem.* **7**, 103–119.
21. Kong, I.-S., Lopez-Casillas, F. & Kim, K.-H. (1990) *J. Biol. Chem.* **265**, 13695–13701.
22. Fry, D. C., Kuby, S. A. & Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 907–911.
23. Hamada, M., Palmieri, R. H., Russell, G. A. & Kuby, S. A. (1979) *Arch. Biochem. Biophys.* **195**, 155–177.
24. Bai, D.-H., Moon, T.-W., Lopez-Casillas, F., Andrews, P. C. & Kim, K.-H. (1989) *Eur. J. Biochem.* **182**, 239–245.
25. Suissa, M., Suda, K. & Schatz, G. (1984) *EMBO J.* **3**, 1773–1781.
26. Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158**, 111–152.
27. Lamhonwah, A.-M., Barankiewicz, T. J., Willard, J. F., Mahuran, D. J., Quan, F. & Gravel, R. Y. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8049–8053.
28. Bowers, D. F. & Allred, J. B. (1991) *FASEB J.* **5**, A426 (abstr.).